NeuroImage 124 (2016) 1196–1201

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Northwestern University schizophrenia data sharing for SchizConnect: A longitudinal dataset for large-scale integration

Alex Kogan [a,*], Kathryn Alpert [a], Jose Luis Ambite [b,c,d], Daniel S. Marcus [e], Lei Wang [a,f]

[a] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[b] Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA
[c] Digital Government Research Center, Marina del Rey, CA, USA
[d] Department of Computer Science, University of Southern California, Los Angeles, CA, USA
[e] Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[f] Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

ARTICLE INFO

Article history:
Accepted 2 June 2015
Available online 16 June 2015

Keywords:
Neuroinformatics
Schizophrenia data
Research datasets
Data mediation and integration
Data sharing
XNAT

ABSTRACT

In this paper, we describe an instance of the Northwestern University Schizophrenia Data and Software Tool (NUSDAST), a schizophrenia-related dataset hosted at XNAT Central, and the SchizConnect data portal used for accessing and sharing the dataset. NUSDAST was built and extended upon existing, standard schemas available for data sharing on XNAT Central (http://central.xnat.org/). With the creation of SchizConnect, we were able to link NUSDAST to other neuroimaging data sources and create a powerful, federated neuroimaging resource.

Introduction

Schizophrenia is a complex disease with heterogeneous clinical, behavioral, cognitive and genetic manifestations, and sharing of datasets is becoming essential in order to test hypotheses that can capture its variability and complexity (Poline et al., 2012). One example is the discovery of microRNA137 that succinctly illustrates the importance of data sharing: using computational biology techniques, Potkin et al. (2010) combined two previously published, separate datasets and discovered microRNA137 as a risk factor for schizophrenia. It should be noted that neither of the two distinct datasets had identified microRNA137. In a later confirmatory report on 51,695 individuals confirming microRNA137, the International Schizophrenia Consortium proclaimed that a new "cause" of schizophrenia had been found (Ripke et al., 2011).

In this paper, we describe the Northwestern University Schizophrenia Data (NUSDAST) (Wang et al., 2013) as part of SchizConnect, an NIH-funded neuroimaging resource for large-scale data sharing for schizophrenia research. With 451 subjects, the majority of whom have archived longitudinal data, NUSDAST is one of the largest single-site, single-platform neuroimaging datasets related to schizophrenia, making it a uniquely important resource to share with the research community. NUSDAST will benefit the neuroscience community in many ways. First, scientists will be able to use these data to generate or test new hypotheses related to abnormalities of brain structures and neural networks in individuals with schizophrenia. Second, scientists will be able to rapidly replicate findings produced using their own datasets. Third, the data could be used to test and validate new brain mapping tools.

What is available?

The data presented in NUSDAST were collected through the support of two NIH-funded grants on schizophrenia: (1) Neuromorphometry in Schizophrenia (R01-MH056584), and (2) Conte Center for the Neuroscience of Mental Disorders (P50 MH071616). Through these projects, our group has collected high-resolution structural MRI datasets from large cohorts of subjects using the same scanner platform and sequence protocols. We have also collected detailed clinical, cognitive and genetic information from these subjects.

Subjects

NUSDAST includes de-identified data from 451 individuals with schizophrenia, their non-psychotic siblings, comparison subjects and their siblings. Neuroimaging data exist for 368 individuals. Longitudinal neuroimaging data are also available on 171 individuals with schizophrenia (m/f = 114/57, age at baseline = 33.8 ± 12.5 years) and 170

* Corresponding author at: Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, 710 N. Lake Shore Dr, Abbott Hall 1322, Chicago, IL 60611.
E-mail address: a-kogan@northwestern.edu (A. Kogan).

http://dx.doi.org/10.1016/j.neuroimage.2015.06.030
1053-8119/

controls (m/f = 86/84, age at baseline = 31.4 ± 13.8 years). Within this group of subjects, 18 individuals with schizophrenia and 30 controls returned for a second follow-up (i.e., 3 time points). The average (SD) follow-up interval was 2.19 (0.82) years for individuals with schizophrenia and 2.28 (0.49) years for the controls. De-identification consisted of stripping HIPAA-mandated identifiable information in research data (such as name, initials, and phone numbers, etc.). Procedures to further anonymize imaging data such as defacing were not performed in order to share the same imaging data that we used in our publications so that others can replicate our findings using their own algorithms if they so desire. See Table 1 below for baseline information.

Clinical data includes information based on specific criteria for clinical stability (Rastogi-Cruz and Csernansky, 1997) and clinical rating scales such as the Scale for the Assessment of Positive Symptoms (Andreasen, 1984) and Scale for the Assessment of Negative Symptoms (Andreasen, 1983) (see Table 1 below for baseline information). Domains of psychopathology (i.e., psychotic symptoms, disorganized symptoms, and negative symptoms) (Andreasen et al., 1995) based on raw scales are also included. The reliability and practicality of using these scales in large populations of schizophrenic patients have been demonstrated by Andreasen, et al. (1995). Symptom assessments were performed by personnel specially trained for this purpose. Inter-rater reliability was monitored regularly for all rating scales, and rater training sessions, including the conjoint assessment of difficult cases, were held weekly. In these sessions, a variety of patients were interviewed in a group. Two established raters reached a consensus of item scores after the interview was completed, and then this “gold standard” score was compared with the rest of the group. New raters were trained by first participating in a minimum of six of these sessions. They were allowed to participate in ratings only after they had demonstrated satisfactory agreement with trained personnel.

MRI data

All MR scans were collected using the same 1.5 T Vision scanner platform (Siemens Medical Systems) at each time point. The Vision scanner had actively shielded gradients and echo-planar capability with very high gradient linearity (<0.4% over a 22-cm diameter spherical volume compared to 2%–5% over 22-cm for our other scanners), which yielded anatomical images with virtually no distortion (<0.4% voxel displacement), critical to analyses of neuroanatomical structures. Using the same scanner provided stable longitudinal MR data throughout the entire period of data collection from 1998 to 2006.

Acquisition of all scans was performed at the Mallinckrodt Institute of Radiology at Washington University School of Medicine, where scanner stability (e.g., frequency, receiver gain, transmitter voltage, SNR) and artifacts were regularly monitored. Phantoms of known size were scanned to confirm image dimensions. Further tests and adjustments (shims, gradient calibrations, EPI switch delays, etc.) were made as needed. During each scan session, a small standardization object (i.e., vitamin E gelcap) was placed on the left side of the forehead for each subject to clearly indicate laterality in the scans. Each scan session included a high-resolution T1-weighted turbo-FLASH scan (Venkatesan and Haacke, 1997), multiple (2–4) MPRAGE scans, and MPRAGE average. Source MR scan data were in Siemens MAGNETOM VISION IMA format and subsequently converted into Analyze format using in-house software. Since Analyze format images may cause confusion with regard to laterality, even though the abovementioned vitamin E gelcap information may help verify laterality, all Analyze format images are being converted into NIFTI format and uploaded. The multiple MPRAGE images for each subject are aligned with the first image and averaged to create a low-noise image volume (Buckner et al., 2004. See Table 2 for detailed scan protocol parameters.

Neuroimaging metadata

In our template-based brain mapping applications, we have focused on a network of structures previously implicated in the pathophysiology of schizophrenia (Weinberger et al., 1992; Csernansky and Bardgett, 1998; Goldman-Rakic, 1999). This network included regions with the prefrontal cortex (e.g., middle frontal gyrus, Brodman area 46) (John et al., 2006; Harms et al., 2010), the cingulate gyrus (Qiu et al., 2007; Wang et al., 2007a), the hippocampus (Wang et al., 2001; Csernansky et al., 2002), the parahippocampal gyrus (Karnik-Henry et al., 2012), as well as the thalamus (Csernansky et al., 2004a; Harms et al., 2007; Smith et al., 2011) and the basal ganglia (Mamah et al., 2007; Wang et al., 2008), which directly or indirectly link these structures via cortical–subcortical connections. We have constructed manual segmentation datasets for all these structures, which can be used for the validation of new computational methods. In addition, we have also used FreeSurfer (Desikan et al., 2006) to generate cortical surface parcellations and measures of cortical regional volume, thickness and surface area (Cobia et al., 2011).

Template data

The templates for the hippocampus and amygdala were generated using a T1-weighted MR scan collected in a healthy subject (Wang et al., 2008). The templates for the thalamus and basal ganglia (caudate nucleus, putamen, nucleus accumbens and globus pallidus) were generated using a seven-time averaged T1-weighted MR scan collected in another healthy subject (Wang et al., 2007b). These segmentations were manually performed using Analyze software in these scans by consensus of experts using atlas guidelines (Duvernoy, 1988, 1991; Mai et al., 1997). Surfaces (.byu format) of each structure were generated using the marching cubes algorithm (Lorensen and Cline, 1987; Claudio and Roberto, 1994). The left and right surfaces have corresponding nodes so that analyses of shape asymmetry can be performed. These templates and subject-level landmark and surface data (below) have been shared here for the purpose of replication and facilitating potential, further modeling work (Haller et al., 1997; Csernansky et al., 2004a; Wang et al., 2007b.

Landmark and surface data

Mapping of the template MR scan occurred in a two-step process. First, it was coarsely aligned to each target scan using landmarks, and then the diffeomorphic map was applied. Surfaces for subcortical structures in the target scans were generated by carrying the template surfaces through these maps (Joshi et al., 1997; Csernansky et al., 2004b.

To facilitate our template-based mapping, global and local (i.e., structure-dependent) neuroanatomical landmarks were placed

Table 1
Subject characteristics at baseline.

	Schizophrenia Subjects	Control Subjects	Schizophrenia Siblings	Control Siblings
N	171	170	44	66
Age at baseline (years)	33.8(12.5 [17–63])	31.4(13.8 [13–67])	N/A	N/A
Gender (male/female)	114/57	86/84	21/23	16/50
Race (Caucasian/African-American/other)	90/78/3	61/105/2	17/27/0	16/50/0
Global SAPS Score	11.1 (12.7 [0–81])	0.06(0.3 [0–4])	0.5 (1.3 [0–7])	N/A
Global SANS Score	9.6 (10.7 [0–62])	0.04(1.7 [0–19])	2.3 (4.8 [0–38])	N/A

Table 2
MR scan parameters.

Sequence	Protocol Parameters
3D turbo-FLASH	TR = 20 ms, TE = 5.4 ms, flip = 30°, ACQ = 1, 256 × 256 matrix, 1 × l mm in-plane resolution, 180 slices, slice thickness 1 mm, 13:30 min scan time
3D MPRAGE (2–4 repeats)	TR = 9.7 ms, TE = 4 ms, flip = 10°, ACQ = 1, 256 × 256 matrix, 1 × l mm in-plane resolution, 128 slices, slice thickness 1.25 mm, 5:36 min scan time each

on the MR images. Landmark-based registration (Joshi et al., 1995) served to adjust the orientation and size for the head (based on global landmarks) and the subcortical structures of interest (based on local landmarks). *Global landmarks*: In each scan, twelve global landmarks were placed following procedures described in Haller et al. (1997) at the points where the anterior and posterior commissures intersected the midsaggittal plane, and at the external boundaries of the cerebrum (anterior, posterior, superior, inferior and lateral). *Local landmarks*: (1) Hippocampus and amygdala were landmarked separately as follows. The most anterior and posterior boundaries of the structure were identified first and a line connecting these points created an anterior/posterior axis. Then, in each of five equally distanced slices along this axis, four landmarks were placed at predetermined points in each slice. (2) Thalamus and basal ganglia were landmarked together as follows. The most anterior boundary of the caudate nucleus and the most posterior boundary of the thalamus were identified and a line connecting these points created an anterior/posterior axis. The region between the two points was then divided into five equally distanced slices along this axis and in each slice five landmarks were placed at predetermined places.

FreeSurfer data

All scans were processed through FreeSurfer Version 3.0.4 (Desikan et al., 2006) pipeline, with careful quality assurance as per FreeSurfer recommendations. All FreeSurfer data, including subcortical segmentation, cortical parcellation and surface, and regional measurement data have been made available.

Cognitive data

Schizophrenia subjects demonstrate a wide array of cognitive deficits (Gur et al., 2001). Data related to intellect, executive functioning (verbal and visual abstraction), and attention, as well as working and episodic memory are included in this dataset. Measures of episodic memory included verbal and visual learning, and also the spontaneous and guided use of memory cues (Jacoby et al., 1993a,1993b, 2001; Jacoby, 1999). Our assessment of working memory included maintenance and manipulation processes across both verbal and visual modalities (Braver et al., 1997; Barch et al., 2002). To date, our schizophrenia subjects have demonstrated deficits across all predicted cognitive domains using this battery (Delawalla et al., 2006; Cobia et al., 2011).

At each visit, the subjects were administered a core battery of neuropsychological measures relevant to areas identified in prior studies of cognition in schizophrenia (Nuechterlein et al., 2004). Tasks were grouped into the following four domains:

Crystallized intelligence

Scaled scores from the vocabulary subtest of the Wechsler Adult Intelligence Scale (WMS-III; (Wechsler, 1997)).

Working memory

Scaled scores based on subtests from the Wechsler Memory Scale – Third Edition (WMS-III; (Wechsler, 1997) including digit span (total forward and backwards), spatial span (total forward and backwards), and letter-number sequencing, and also overall d-prime from the CPT-IP task (Cornblatt et al., 1988).

Episodic memory

Included scaled scores from the WMS-III Logical Memory and Family Pictures subtests.

Executive function

Included number of novel words generated on phonemic and semantic verbal fluency tasks (Benton, 1976; Benton et al., 1984), time to completion on the Trail Making Test Part B (Reitan and Wolfson, 1985), scaled scores on the WAIS-III Matrix Reasoning subtest, and number of perseverative errors on the Wisconsin Card Sorting Test (Heaton et al., 1993).

Cronbach's alpha (assessed in the standardization set of subjects) was 0.77, 0.78 and 0.70 for working memory, episodic memory and executive function, respectively, in individuals with schizophrenia, and 0.76, 0.65 and 0.67, respectively, in control individuals.

Genotyping data

Blood for the isolation of DNA was collected from each of the subjects. These samples have been genotyped for a panel of 20 gene polymorphisms selected for their association with schizophrenia or their involvement in neurodevelopment. Examples of these genes include BDNF (rs6265), EGFR (rs10228436), FGF-2 (rs1048201) and IL-3 (rs40401). Morphometric measures (e.g., structural volume) of individuals are compared and contrasted with specific differences (i.e., single nucleotide polymorphisms or SNPs) in the genes of interest. These differences are qualified by testing whether or not each subject has a particular polymorphism, and then how many copies of that polymorphism they have. A subject can fall into one of three categories: both copies of the gene are polymorphism-free (homozygous), one copy is polymorphism-free whereas the other copy has the polymorphism (heterozygous), or both copies of the gene carry the polymorphism (homozygous).

Presently, we have genotyping data on 117 subjects with schizophrenia and 58 controls. DNA samples in additional subjects are being analyzed. All available genotype data will be made available with the scans to users of the database.

Data dictionary

Along with the data, we provide a data dictionary of terms. In the dictionary, standard descriptions for which ontologies exist are used. We searched the following sources for ontology: NeuroLEX (http://neurolex.org/wiki/Main_Page), a semantic wiki for terms used in neuroscience; the Neuroscience Information Framework (http://www.neuinfo.org/), a dynamic resource of web-based neuroscience data, materials, and tools (NeuroLEX terms are actually published in NIF); and NCI Metathesaurus http://ncimeta.nci.nih.gov/ncimbrowser/), a biomedical terminology database for translational and basic research. A detailed list of these terms is presented in Table 3, and examples include "socioeconomic status," "SAPS" and "cognitive assessment." For descriptions of without a standard ontology, such as "working memory" or "global rating of hallucinations," we plan to work with NeuroLEX to arrive at standard definitions. The current version of the data dictionary can be downloaded through the data portal website, described below.

Data sharing mechanisms

Data sharing architecture: XNAT and XNAT Central

The collected MR datasets along with detailed clinical information are archived using the eXtensible Neuroimaging Archive Toolkit (XNAT), an open source data management and productivity platform for biomedical imaging research. XNAT was developed by the Neuroinformatics Research Group (NRG) at Washington University in St. Louis and the BIRN (Marcus et al., 2007a,2007b). It is widely used

Table 3
Table of available ontology for data dictionary terms.

Ontology Term	NeuroLex Name & ID NIF Standard Ontology ID NCI Methathesaurus ID Cognitive Atlas
Gender	NeuroLex: Gender assessment, birnlex_3026; NIF: nif_inv:birnlex_3026; NCI: C44177
Race	NeuroLex: Race assessment, birnlex_3040; NIF: nif_inv:birnlex_3040; NCI: C17049
Group	NeuroLex: Control role, birnlex_11017[a]
Ethnicity	NeuroLex: Ethnicity assessment, birnlex_3015; NIF: nif_inv:birnlex_3015; NCI: C16564
Marital status	NeuroLex: Marital status assessment, birnlex_3031; NIF: nif_inv:birnlex_3031; NCI: C25188
Type of housing/living arrangement	NeuroLex: Living arrangement assessment, birnlex_3030; NIF: nif_inv:birnlex_3030; NCI: C94852
Number of siblings	NCI: C102469
Number of children	NeuroLex: Offspring cardinality assessment, birnlex_3035; NIF: nif_inv:birnlex_3035
Current occupation/job title	NeuroLex: Occupation assessment, birnlex_3036[b]; NIF: nif_inv:birnlex_3036; NCI: C25193
Principal occupation	NeuroLex: Occupation assessment, birnlex_3036[b]; NIF: nif_inv:birnlex_3036; NCI: C25193
Level of education	NeuroLex: Education assessment, birnlex_3014[c]; NIF: nif_inv:birnlex_3014
Years of schooling	NeuroLex: Education assessment, birnlex_3014[c]; NIF: nif_inv:birnlex_3014; NCI: C17953
Father's level of education	NeuroLex: Father's education, birnlex_3021[d]; NIF: nif_inv:birnlex_3021
Father's years of schooling	NeuroLex: Father's education, birnlex_3021[d]; NIF: nif_inv:birnlex_3021
Mother's level of education	NeuroLex: Mother's education, birnlex_3023[e]; NIF: nif_inv:birnlex_3023
Mother's years of schooling	NeuroLex: Mother's education, birnlex_3023[e]; NIF: nif_inv:birnlex_3023
Socioeconomic status	NeuroLex: Socio-Economic status, birnlex_3048 NIF: nif_inv:birnlex_3048 NCI: C17468
Handedness	NCI: (CUI) C0023114
Edinburg handedness assessment	NeuroLex: Edinburg handedness assessment, birnlex_3013; NIF: nif_inv:birnlex_3013
SAPS	NeuroLex: Scale for the Assessment of Positive Symptoms, birnlex_3045[f]; NIF: nif_inv:birnlex_3045
SANS	NeuroLex: Scale for the Assessment of Negative Symptoms, birnlex_3041[g]; NIF: nif_inv:birnlex_3041
Cognitive assessment	NeuroLex: birnlex_2021; NIF: nif_inv:birnlex_2021; NCI: C0870300
Crystallized intelligence	CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d09f64
Working memory	CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0b5a7
Episodic memory	CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0a1f4
Executive function	CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0a252

[a] Currently a proxy class to be replaced by its OBI (The Ontology for Biomedical Investigations) equivalent.
[b] NeuroLex Occupation assessment includes work specialties as defined by duties and required skills as well as principal activity that a person does to earn money.
[c] NeuroLex Educational assessment includes level of education and years of schooling.
[d] NeuroLex Father's Education assessment includes father's level of education and father's years of schooling.
[e] NeuroLex Mother's Education assessment includes mother's level of education and mother's years of schooling.
[f] NeuroLex SAPS describes a type of assessment without any defined attributes.
[g] NeuroLex SANS describes a type of assessment without any defined attributes.

across the world and is a core component of the emerging NIH-backed biomedical informatics backbone, including the Biomedical Informatics Research Network (BIRN) and the National Alliance for Medical Image Computing (NA-MIC). XNAT includes a secure database backend and a rich web-based user interface. See paper on XNAT Central in this special issue for more details.

XNAT custom schemas

XNAT relies on extensible markup language (XML) schema documents to define the type of data that can be stored in the system. XML is the standard language for defining open and extensible data formats. The XML format provides a number of benefits for data organization: it uniformly describes data and data structure, it makes data available for consistent and efficient programmatic manipulation, reuse, transmission and storage, and it simplifies data conversion to other formats. XNAT comes with a set of XML schemas that describe common data associated with neuroimaging studies. XNAT also allows for the extension of these schemas as well as the creation of custom schemas. NUSDAST contains all three types: common, extended and custom schemas. The extended and custom schemas include subject registration data and extended demographic and relationship information.

Data access

XNAT Central

The project "NU Schizophrenia Data and Software Tool Federation using BIRN Infrastructure (NUSDAST)" is hosted on XNAT Central (visit http://central.xnat.org, keyword NUSDAST or directly at http://tinyurl.com/av9h7jm).[1] Within the NUSDAST project on the XNAT Central website, data are organized by subject following the XNAT architecture: study registration and longitudinal epochs of MR sessions. Within each epoch's MR session, scans and associated segmentations, surfaces, landmarks and other data are listed. User download is accomplished via the Download action or the Manage Files action displayed on the XNAT web page. Data can also be retrieved via the XNAT REST API. Only users with validated accounts are able to query and download data in NUSDAST.

Web-based SchizConnect portal

Our recent effort to create a new and innovative way to access, query and retrieve neuroimaging data from various distributed research databases resulted in a web-based portal called SchizConnect (Wang et al., 2014) (http://www.schizconnect.org). NUSDAST is one of the sources integrated with SchizConnect. SchizConnect communicates with NUSDAST via XNAT REST API in order to search and retrieve data based on criteria entered on the SchizConnect web interface. With SchizConnect, users can get a summary of the NUSDAST data made available to them. In order to be able to retrieve NUSDAST data through SchizConnect, users are required to sign a data usage agreement (DUA). DUAs are used to restrict the use and disclosure of the available data as well as to permit publication of the research results in accordance with the applicable laws and regulations. The DUAs for SchizConnect and the current sources (COINS, NUSDAST, MCIC and FBIRN) require users to acknowledge the source of the data as well as the funding source in any publications and presentations, to protect the privacy of the subjects the data was collected from, and to keep data from being transferred to any third-party users that have not signed the DUAs related to the dataset. See paper on SchizConnect in this special issue for more details.

The significant advantage of using SchizConnect is that it allows the user to combine neuroimaging data from different databases to create a mediated dataset with related data. In our case, cognitive data, psychopathology measures based on SAPS and SANS and SNP data is stored at a different research database called REDCap. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing (1) an intuitive interface for validated data entry; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for importing data from external sources. (Harris et al., 2009). Since cognitive, SAPS/SANS and SNP data stored in the REDCap source are related to the neuroimaging data stored in NUSDAST, SchizConnect will seamlessly provide the user with all of the available data in one dataset, even though the data comes from geographically and technologically different sources.

[1] Full URL: https://central.xnat.org/app/template/XDATScreen_report_xnat_projectData.vm/search_element/xnat:projectData/search_field/xnat:projectData.ID/search_value/NUDataSharing.

Discussion

In this paper, we described an instance of the Northwestern University Schizophrenia Data (NUSDAST), a static, longitudinal schizophrenia-related dataset, along with the XNAT Central hosting platform and the SchizConnect data portal used for accessing and sharing the dataset. All the clinical and longitudinal data was collected during the performance of two NIH-funded studies and all the data constituting the dataset was reviewed and approved by the investigators of these studies.

Concerning data sharing, this resource built and extended upon existing, standard schemas available for data sharing on XNAT Central (http://central.xnat.org/). Specifically, we developed additional schemas for storing demographic metadata in XNAT. This addition creates the opportunity to consistently expand and share schizophrenia research–related data.[2] We have significantly improved the way scientists are able to mine our dataset by integrating with the SchizConnect web portal for searching and downloading our data along with the accompanying longitudinal data.

Our well-described and comprehensive data on the normal controls and their siblings are valuable beyond the schizophrenia research community. For example, the SNP data, located at the REDCap resource, includes ones that are related to neurodevelopment (e.g., BDNF), embryonic development and tissue repair (e.g., FGF-2), and immune response/inflammation (e.g., EGFR, IL-3). Mutations of many of these SNPs have been found to be related to cancer and neuropsychiatric disorders such as depression, anxiety and Alzheimer's disease. Therefore, with the accompanying imaging and cognitive data, our control subjects' data can be of wider use, beyond schizophrenia research.

The schizophrenia research community has invested substantial resources in order to collect, manage and share increasingly larger datasets including neuroimaging data. The exploration of large, multimodal datasets has indeed improved our understanding of relationships among abnormalities of brain circuitry, brain function and genetic variability in schizophrenia (Kim et al., 2009; Kim et al., 2010; Allen et al., 2011.

Numerous data sharing initiatives were undertaken in order to create publicly accessible neuroimaging data collections, such as FBIRN, MCIC, BSNIP, fMRI DataCenter (fMRIDC) and the Open Access Series of Imaging Studies (OASIS). One of the main obstacles to the open access sharing of research data is the lack of local organization and standard descriptions (Poline et al., 2012). Different resources usually organize data in many different formats, which leads to difficulties in sharing and analyzing data. For example, the acquisition of a dataset from any given source will most likely require some programming work in order to make the dataset suitable for processing by an already utilized suite of tools. The introduction of the SchizConnect data portal implements the concept of "one-click share." The SchizConnect tool harmonizes different research databases in order to create a mediated mega-dataset with related longitudinal data. SchizConnect is a step in the direction of the creation of standardized procedures that include proliferation of open-source data sharing and of storage mechanisms. The SchizConnect portal responds to the current needs of researchers as it is a comprehensive and extensible system that accommodates different types of research and provides data-mining for common data processing functionality.

It is our hope that this effort will help overcome some of the commonly recognized technical barriers to advancing neuroimaging research (Poline et al., 2012) by creating a research-ready dataset that

[2] Extended and new schemas created by NUSDAST can be used by other existing or future projects to describe their data. A benefit of sharing data through XNAT Central, rather than through an independent XNAT system, is that multiple projects can easily use the same schema representations and user interface components to represent common data elements. Users can then query and mine XNAT Central to locate data across the multiple projects using the shared schema.

meaningfully combines neuroimaging data with other relevant information. Currently, more data are being made available, such as fMRI and genome-wide scan (GWS) data. We are expanding the scope of schizophrenia neuroimaging data sharing by linking NUSDAST with FBIRN and MCIC through the development of SchizConnect, which is a powerful data mediation and integration platform that establishes a true federation of disparate, heterogeneous neuroimaging-related databases.

Acknowledgment

This work was supported in part by NIH grants 1R01 MH084803, 1U01 MH097435-01A1 and R01 EB009352.

References

Allen, E.A., Erhardt, E.B., et al., 2011. A baseline for the multivariate comparison of resting-state networks. Front. Syst. Neurosci. 5, 2.
Andreasen, N.C., 1983. The Scale for Assessment of Negative Symptoms (SANS). The University of Iowa, Iowa City, Iowa.
Andreasen, N.C., 1984. The Scale for Assessment of Positive Symptoms (SAPS). The University of Iowa, Iowa City, Iowa.
Andreasen, N.C., Arndt, S., et al., 1995. Symptoms of schizophrenia. Methods, meanings, and mechanisms. Arch. Gen. Psychiatry 52, 341–351.
Barch, D.M., Csernansky, J., et al., 2002. Working and long-term memory deficits in schizophrenia. Is there a common underlying prefrontal mechanism? J. Abnorm. Psychol. 111, 478–494.
Benton, A.L., 1976. Multilingual Aphasia Examination. University of Iowa, Iowa City.
Benton, A.L., Hamsher, K.d.S., et al., 1984. Multilingual Aphasia Examination. Third Edition. Psychological Assessment Resources, Iowa City (MAE).
Braver, T.S., Cohen, J.D., et al., 1997. A parametric study of prefrontal cortex involvement in human working memory. NeuroImage 5 (1), 49–62.
Buckner, R.L., Head, D., et al., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. NeuroImage 23 (2), 724–738.
Claudio, M., Roberto, S., 1994. Using Marching Cubes on small machines. Graph. Model. Image Proc. 56, 182–183.
Cobia, D.J., Csernansky, J.G., et al., 2011. Cortical thickness in neuropsychologically near-normal schizophrenia. Schizophr. Res. 133 (1-3), 68–76.
Cornblatt, B.A., Risch, N.J., et al., 1988. The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. Psychiatry Res. 26 (2), 223–238.
Csernansky, J.G., Bardgett, M.E., 1998. Limbic-cortical neuronal damage and the pathophysiology of schizophrenia. Schizophr. Bull. 24 (2), 231–248.
Csernansky, J.G., Wang, L., et al., 2002. Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping. Am. J. Psychiatry 159 (12), 2000–2006.
Csernansky, J.G., Schindler, M.K., et al., 2004a. Abnormalities of thalamic volume and shape in schizophrenia. Am. J. Psychiatry 161 (5), 896–902.
Csernansky, J.G., Wang, L., et al., 2004b. Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. NeuroImage 23 (Suppl. 1), S56–S68.
Delawalla, Z., Barch, D.M., et al., 2006. Factors mediating cognitive deficits and psychopathology among siblings of individuals with schizophrenia. Schizophr. Bull. 32 (3), 525–537.
Desikan, R.S., Segonne, F., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (3), 968–980.
Duvernoy, H.M., 1988. The Human Hippocampus: An Atlas of Applied Anatomy. Munich, J. F. Bergmann Verlag.
Duvernoy, H.M., 1991. The Human Brain: Surface, Three-Dimensional Anatomy and MRI. Springer-Verlag, Wien, New York.
Goldman-Rakic, P.S., 1999. The physiological approach: Functional architecture of working memory and disordered cognition in schizophrenia. Biol. Psychiatry 46 (5), 650–661.
Gur, R.C., Ragland, J.D., et al., 2001. Computerized neurocognitive scanning: II. The profile of schizophrenia. Neuropsychopharmacology 25 (5), 777–788.
Haller, J.W., Banerjee, A., et al., 1997. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology 202 (2), 504–510.
Harms, M.P., Wang, L., et al., 2007. Thalamic shape abnormalities in individuals with schizophrenia and their nonpsychotic siblings. J. Neurosci. 27 (50), 13835–13842.
Harms, M.P., Wang, L., et al., 2010. Structural abnormalities in gyri of the prefrontal cortex in individuals with schizophrenia and their unaffected siblings. Br. J. Psychiatry 196 (2), 150–157.
Harris, P.A., Taylor, R., et al., 2009. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42 (2), 377–381.
Heaton, R.K., Chelune, G.J., et al., 1993. Wisconsin Card Sorting Test manual. Psychological Assessment Resources, Inc., Odessa, Florida.
Jacoby, L.L., 1999. Deceiving the elderly: Effects of accessibility bias in cued-recall performance. Cogn. Neuropsychol. 16, 001–020 (000-000).
Jacoby, L.L., Ste-Marie, D., et al., 1993a. Redefining automaticity: Unconscious influences, awareness, and control. In: Baddeley, A., Weiskrantz, L. (Eds.), Attention: Selection, Awareness, and Control; A Tribute to Donald Broadbent. Clarendon, New York, pp. 261–282.
Jacoby, L.L., Toth, J.P., et al., 1993b. Separating conscious and unconscious influences of memory: Measuring recollection. J. Exp. Psychol. Gen. 122 (2), 139–154.
Jacoby, L.L., Debner, J.A., et al., 2001. Proactive interference, accessability bias, and process dissociations: Valid subjective reports of memory. J. Exp. Psychol. Learn. Mem. Cogn. 27, 686–700.
John, J.P., Wang, L., et al., 2006. Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri. Psychiatry Res. 148 (2-3), 151–163.
Joshi, S., Miller, M.I., et al., 1995. The generalized Dirichlet problem for mapping brain manifolds. Int'l Symp on Optical Science, Engineering, and Instrumentation, Vision Geometry IV, SPIE.
Joshi, S., Miller, M.I., et al., 1997. On the geometry and shape of brain sub-manifolds. Int. J. Pattern Recognit. Artif. Intell. 11 (8), 1317–1343 (Special Issue).
Karnik-Henry, M.S., Wang, L., et al., 2012. Medial temporal lobe structure and cognition in individuals with schizophrenia and in their non-psychotic siblings. Schizophr. Res. 138 (2-3), 128–135.
Kim, D.I., Manoach, D.S., et al., 2009. Dysregulation of working memory and default-mode networks in schizophrenia using independent component analysis, an fBIRN and MCIC study. Hum. Brain Mapp. 30 (11), 3795–3811.
Kim, M.A., Tura, E., et al., 2010. Working memory circuitry in schizophrenia shows widespread cortical inefficiency and compensation. Schizophr. Res. 117 (1), 42–51.
Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3D surface construction algorithm. Comput. Graph. 21 (4), 163–169.
Mai, J.K., Assheuer, J., et al., 1997. Atlas of the Human Brain. Academic Press, San Diego.
Mamah, D., Wang, L., et al., 2007. Structural analysis of the basal ganglia in schizophrenia. Schizophr. Res. 89 (1-3), 59–71.
Marcus, D.S., Olsen, T.R., et al., 2007a. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics 5 (1), 11–34.
Marcus, D.S., Wang, T.H., et al., 2007b. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19 (9), 1498–1507.
Nuechterlein, K.H., Barch, D.M., et al., 2004. Identification of separable cognitive factors in schizophrenia. Schizophr. Res. 72 (1), 29–39.
Poline, J.B., Breeze, J.L., et al., 2012. Data sharing in neuroimaging research. Front. Neuroinformatics 6, 9.
Potkin, S.G., Macciardi, F., et al., 2010. Identifying gene regulatory networks in schizophrenia. NeuroImage 53 (3), 839–847.
Qiu, A., Younes, L., et al., 2007. Combining anatomical manifold information via diffeomorphic metric mappings for studying cortical thinning of the cingulate gyrus in schizophrenia. NeuroImage 37 (3), 821–833.
Rastogi-Cruz, D., Csernansky, J., 1997. Clinical Rating Scales. Adult Psychiatry. Mosby, Inc., S. Guze. St. Louis.
Reitan, R.M., Wolfson, D., 1985. The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation. Neuropsychology Press, Tucson, Arizona.
Ripke, S., Sanders, A.R., et al., 2011. Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. 43 (10), 969–976.
Smith, M.J., Wang, L., et al., 2011. Thalamic morphology in schizophrenia and schizoaffective disorder. J. Psychiatr. Res. 45 (3), 378–385.
Venkatesan, R., Haacke, E., 1997. Role of high resolution in magnetic resonance (MR) imaging: Applications for MR angiography, intracranial T1-weighted imaging, and image interpolation. Int. J. Imaging Syst. Technol. 8, 529–543.
Wang, L., Joshi, S.C., et al., 2001. Statistical analysis of hippocampal asymmetry in schizophrenia. NeuroImage 14 (3), 531–545.
Wang, L., Hosakere, M., et al., 2007a. Abnormalities of cingulate gyrus neuroanatomy in schizophrenia. Schizophr. Res. 93 (1-3), 66–78.
Wang, L., Lee, D.Y., et al., 2007b. Validity of large-deformation high dimensional brain mapping of the basal ganglia in adults with Tourette syndrome. Psychiatry Res. 154 (2), 181–190.
Wang, L., Mamah, D., et al., 2008. Progressive deformation of deep brain nuclei and hippocampal-amygdala formation in schizophrenia. Biol. Psychiatry 64 (12), 1060–1068.
Wang, L., Kogan, A., et al., 2013. Northwestern University Schizophrenia Data and Software Tool (NUSDAST). Front. Neuroinformatics 7, 25.
Wang, L., Alpert, K.I., et al., 2014. SchizConnect: Large-Scale Schizophrenia Neuroimaging Data Integration and Sharing. Annual Meeting of the American College of Neuropsychopharmacology (ACNP). Phoenix, Arizona.
Wechsler, D., 1997. Wechsler Adult Intelligence Scale. The Psychological Corporation, San Antonio.
Weinberger, D.R., Berman, K.F., et al., 1992. Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. Am. J. Psychiatry 149 (7), 890–897.